

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Structure and Motion from Image Sequences based on Multi-Scale Bayesian Network

Norio Tagawa and Shoichi Naganuma
*Tokyo Metropolitan University
Japan*

1. Introduction

A lot of studies have been reported on the problem of structure from motion (SFM) as a central theme of computer vision (CV). In the beginning of the research in this field, the principles of 3-D depth recovery and 3-D motion estimation from the viewpoint of mathematics have the attention of a lot of researchers (Adiv, 1985; Huang & Faugeras, 1989; Kanatani, 1993; Longuet-Higgins, 1981; Maybank, 1990; Tagawa et al., 1993; Tsai & Huang, 1984; Zhuang et al., 1988). Subsequently accurate recovery methods have been examined and simultaneously the idea and the role of stochastic inference in CV have been discussed and analyzed (Daniilidis & Nagel, 1990; Kanatani, 1996; Tagawa et al., 1994; Tagawa et al. 1996). Recently, stable and efficient methods represented by the factorization technique (Han & Kanade, 2002; Ke & Kanade, 2005) and methods with no use of camera calibration (Han & Kanade, 2002) have been proposed. Using these methods the development of the virtual reality technique has been intensely advanced. However, the intuitive difficulties of SFM, such as accuracy, high-resolution, computational cost and so on, have not been solved completely. Namely, a practical method for accurately detecting dense motion fields in images and/or relative depth maps between a camera and a target object, keeping spatial discontinuity with low computational cost, has not been established, and many studies have progressed on this problem (Brox et al., 2004; Farnebäck, 2001). Although the method based on the Markov random field (MRF) including a line process (Geman & Geman, 1984) is systematic and the resultant accuracy is significant, its realization in human vision system is not easy because of its computational complexity. On the other hand, increasing the observation information by unifying multiple frames is an important strategy, and the recently regarded methods described above use multi-frame information suitably (Bruhn & Weickert, 2005). However, these techniques assume that tracking of sparse feature points has been performed in advance, and hence, reliable detection caused by integrating temporal information is derived only at the sparse pixels. In this study we introduce a method which can recover dense and accurate depth maps based on two successive frames with no use of the complex MRF including a line process. In this framework, we are going to consider temporal unification in our future research.

Since the detection of 2-dimensional motion field, called optical flow, based on the gradient equations is an ill-posed problem, another condition is required to determine the value of optical flow at each pixel. This issue is called "aperture problem," and it is a fundamental

difficulty causing heteroptics in human vision. In the direct method (Horn & Welden Jr, 1988; Stein & Shashua, 1997) which recovers depth without explicit detection of optical flow, the aperture problem also arises. The aperture problem has been conventionally avoided by either the local optimization (Lucas & Kanade, 1981; Kearney et al., 1987) or the global optimization (Horn & Schunk, 1981). The former assumes that optical flow or depth is locally constant. The latter assumes that the optical flow or depth changes smoothly in the spatial and/or the temporal domains. However, these assumptions impose constraints on the shape of the target directly or indirectly and cause resolution deterioration of the recovered structure. Especially, at the pixels where depth varies discontinuously, for example the contour of the object, the recovered depth might be inaccurate. Lately, in consideration of the case in which intensity invariableness before and after relative camera motion does not hold, methods using intensity constraint as well as geometric constraint are examined (Maki et al. 2002). Although it is possible that these schemes can solve the aperture problem, the research on this issue is still in progress.

As another difficulty with respect to the optical flow detection, the alias problem should be solved. When the intensity pattern with short wavelength in comparison with the size of optical flow is used to detect optical flow, large detection error is observed. This detection error coincides with a usual aliasing phenomenon, and hence a frame rate is not enough to get complete information of the intensity variation. In an active vision scheme, the alias problem can be avoided by making the size of optical flow under a certain value constantly with a suitable time-sampling interval. However, for usual applications, passively taken image sequences are often used to recover depth maps, and then, a method based on a signal processing scheme is desirable. Most of the conventional methods extract spatially smooth intensity patterns by low-pass filtering. Therefore using these techniques, only a low spatial-resolution structure is recovered.

In our study, original images are decomposed into multi-scale images, and the depth information detected using low resolution images is propagated to high resolution images in order to avoid the alias problem and to realize stable and high-resolution recovery using the dynamic Bayesian network spreading to a resolution direction. In this statistical inference processing, by applying the local optimization method and making the local region size smaller as the resolution of the treated images increases, depth discontinuity can be exactly recovered with low computational cost in comparison with the MRF. We introduce an algorithm by which depth map is recovered directly from the spatial and temporal variations of the image intensity without detecting optical flow. Although most of the related research firstly detects optical flow (Tagawa et al., 1995), and recovers depth map using the detected optical flow as an intervening measurement (Tagawa et al., 1996), it is natural for human vision system to extract and recognize optical flow, which is not necessarily caused by rigid motion, and depth map in parallel from the temporal variation of the light intensity received on a retina. From a computational theory, firstly detected optical flow without rigid motion constraints can be regarded as an intermediate solution derived by expanding the solution space, and hence it is not sure that the finally obtained depth map satisfies the exact constraints.

In the dynamic Bayesian network used in this study, each unknown parameter is represented as a node as well as the depth corresponding to each pixel to be recovered and the observed image information. We call this graphical model a multi-scale Bayesian network. If the parameters, including relative 3-D motion between the camera and the object,

are determined in advance, the inference of depth map is realized by the Kalman filter (Huang & Ho, 1999; Matthies & Kanade, 1989). For optical flow detection, Simoncelli proposed a method based on the same network in which optical flow is considered as a node and parameters are assumed to be known (Simoncelli, 1999).

The scheme we propose in this chapter is to estimate depth map and parameters simultaneously from image observations using the above described Bayesian network. The parameters to be estimated are common to all multi-scale images, and hence, complete information propagation of both depth and parameters is complicated. Therefore, a suitable approximation has to be adopted. For that purpose we proposed a method based on the MAP-EM algorithm, in which the Laplace approximation is applied to the posterior probabilities of the parameters, and the saddle point approximation is introduced to the posterior of the depth (Tagawa et al., 2008). However, in this method, the variances of the parameters are computed by naïve numerical evaluation of the second order differentials of the log likelihood, which estimation is unstable and inaccurate. In this chapter, we present a new algorithm, in which the Supplemented MAP-EM algorithm (Meng & Rubin, 1991; van Dyk et al., 1995) is adopted to achieve a stable and accurate estimator of the above variances.

2. Principle of Direct Structure and Motion Recovery

2.1 Optical flow caused by rigid motion

We use perspective projection as our camera-imaging model shown in Fig. 1. The camera is fixed with an (X, Y, Z) coordinate system, where the viewpoint, i.e., lens center, is at origin O and the optical axis is along the Z -axis. The projection plane, i.e. image plane, $Z = 1$ can be used without any loss of generality, which means that the focal length equals 1. A space point (X, Y, Z) on the object is projected to the image point (x, y) . At each (x, y) , the optical flow $[v_x, v_y]^T$ is formulated with an inverse depth $d(x, y) = 1/Z(x, y)$ and the camera's translational and rotational vectors $\mathbf{u} = [u_x, u_y, u_z]^T$, and $\mathbf{r} = [r_x, r_y, r_z]^T$, respectively, are given as follows:

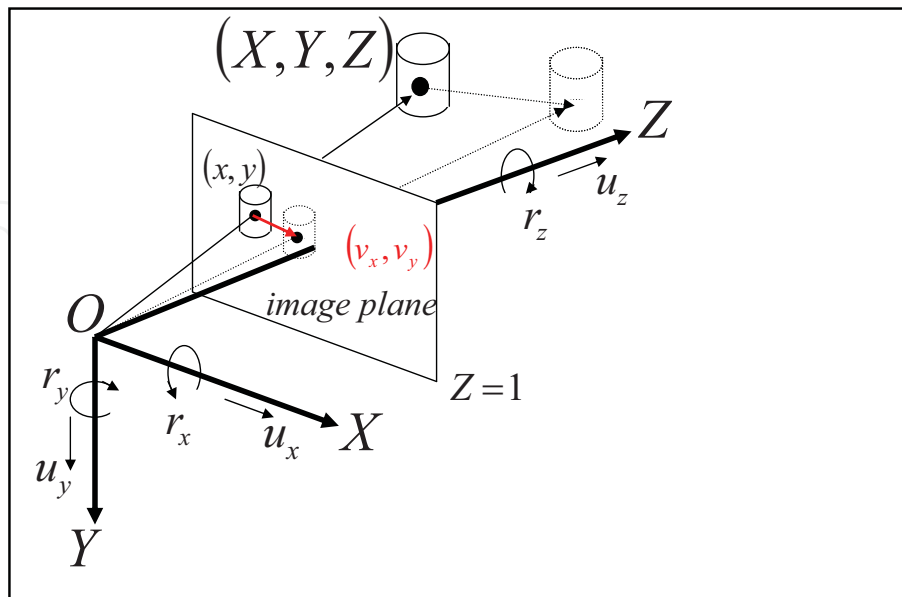


Fig. 1. Camera-imaging model used in this chapter

$$v_x = xy r_x - (1 + x^2) r_y + y r_z - (u_x - x u_z) d, \quad (1)$$

$$v_y = (1 + y^2) r_x - x y r_y - x r_z - (u_y - y u_z) d. \quad (2)$$

In the above equations, d is an unknown variable at each pixel position, and \mathbf{u} and \mathbf{r} are unknown parameters common for the whole image. In the following, Eqs. 1 and 2 are rewritten as

$$v_x = v_x^r(\mathbf{r}) + v_x^u(\mathbf{u})d, \quad (3)$$

$$v_y = v_y^r(\mathbf{r}) + v_y^u(\mathbf{u})d. \quad (4)$$

2.2 Gradient equation for rigid motion

The optical flow constraint equation, which is called the “gradient equation,” is the first approximation of the assumption that image intensity is invariable before and after the relative 3-D motion between a camera and an object. At each pixel (x, y) in the image, the gradient equation is formulated with the partial differentials f_x , f_y and f_t , where t denotes time, of the image intensity $f(x, y, t)$ and the optical flow

$$f_t = -f_x v_x - f_y v_y. \quad (5)$$

By substituting Eqs. 3 and 4 into Eq. 5, the gradient equation representing a rigid motion constraint can be derived explicitly as

$$\begin{aligned} f_t &= -(f_x v_x^r + f_y v_y^r) - (f_x v_x^u + f_y v_y^u) d \\ &\equiv -f^r - f^u d. \end{aligned} \quad (6)$$

This equation can be used for the direct recovering the structure and the motion (Horn & Weldon Jr, 1988; Stein & Sashua, 1997). In this study, we use a convolution kernel defined as a derivative of the suitable interpolator kernel (Farid & Simoncelli, 1997), which is described in Sec.6.2 in detail, to calculate f_x and f_y accurately, and f_t is detected as the finite difference using two successive frames. Hence, we suppose that only f_t contains observation error, and we use Eq. 6 as the observation equation.

It is obvious that from Eq. 6 the norm of \mathbf{u} and d cannot be uniquely determined. Therefore, we suppose that the norm of \mathbf{u} equals 1, and consider \mathbf{u} to be $\mathbf{u} = [u_x, u_y, (1 - u_x^2 - u_y^2)^{1/2}]^T$.

3. Outline of Depth Recovery with Multi-Scale Processing

3.1 Problems definition

In order to recover accurate and high resolution depth map, the alias problem and the aperture problem have to be avoided. Firstly, we briefly explain the aperture problem. By representing the optical flow at each pixel as $[v_x, v_y]^T$, from the relation $k_t = v_x k_x + v_y k_y$ between the spatial wavenumber (k_x, k_y) and the temporal wavenumber k_t , it can be known that the Nyquist frequency of time-sampling takes different value for each spatial frequency.

However, usual time-sampling rate for each pixel, i.e. frame rate, is constant independently of the spatial frequencies. Hence, the possibility that the time-sampling rate for high spatial frequencies can not satisfy the sampling theory is higher than such possibility for low spatial frequencies. Therefore, a lot of conventional methods analyze only the low resolution images extracted by low-pass filtering. In these techniques, high resolution information appears to be lost, which causes deterioration of the recovered structure.

Subsequently, the aperture problem means that if some pairs of (f_x, f_y) take the same values in the local region in the image, where the optical flow can be assumed to be constant, we can not determine the optical flow uniquely by solving simultaneously the multiple equations corresponding to Eq. 5 obtained at the pixels in such local region. The essential cause for the aperture problem is the fact that Eq. 5 is ill-posed. The conventional methods which can cope with the problem are roughly divided into two types: the local optimization method and the global optimization method. The local optimization method assumes that the optical flow or the corresponding depth in each spatial local region and/or in each temporal local region is constant. Such the local regions have to be determined so that (f_x, f_y) s take various values. Then, each optical flow and each corresponding depth are obtained using the multiple equations in Eq. 5. By this approach, local constraints are introduced into the unknown structure, and only linear computations are required. On the other hand, the global optimization method uses the global constraints represented by the spatial smoothness of the optical flow or the depth, and concretely solves differential equations derived through the variational principle.

As mentioned above, introducing limitations to the image information used to recover depth, i.e. low frequency components are extracted and used, for the alias problem, and introducing limitations to the unknown structure, i.e. resolution and hence degree of freedom of structure are lowered, for the aperture problem, are important approaches. Much conventional methods solve the aperture problem and the alias problem by applying such kinds of limitations. However, they cause resolution lowering of the obtained structure as compared with the raw image resolution, and then, accuracy of recovering is decreased especially at the discontinuous parts of the structure.

3.2 Fundamental concept of proposed solution

In the method described in this chapter, we aim to solve the above described two problems without lowering the resolution of recovered structure. Fundamental ideas are (i) decomposing images into multi-scale images each of which has proper spatial frequencies, (ii) adapting the assumed resolution of the structure to the resolution of the used image at each scale processing, and (iii) advancing the processing from low resolutions to high resolutions sequentially.

The combination of (i) and (iii) deals with the alias problem. Although it is desirable that low resolution images are used in order to avoid aliasing, it is important that high resolution images are suitably analyzed in order not to lower the resolution of the recovered structure. Therefore, the employment of high resolution information with avoiding the aliasing can be performed by sequentially propagating the stable depth map with no aliasing obtained from low resolution step to high resolution step. This strategy is the same as the one proposed in the previous study (Simoncelli, 1999) for optical flow detection. By adopting the Bayesian

inference as an information propagation scheme, depth error in low resolution can be compensated by high resolution processing.

The combination of (i) and (ii) deals with the aperture problem. If we adopt the scheme (i) to avoid the aliasing, the obtained multi-scale depth information has naturally a hierarchical structure. Hence, it is appropriate that spatial resolution of the depth variable is lowered for low resolution step and inversely it is improved for high resolution step. Such control of the resolution depending on the image resolution can be done by both the local optimization method and the global optimization method, and in this study, we use the local optimization scheme, since it can be performed by local computations. Basically, for the low resolution step the size of the region where depth is assumed to be constant is expanded, and as resolution becomes high, this size is contracted.

In addition to the methodological feature of our solution described above, the following technical originality is asserted in this study. If the parameters contained in our statistical model are determined in advance, the Kalman filter can be applied to the inference of depth map (Huang & Ho, 1999; Matthies & Kanade, 1989). However, such condition can not be generally supposed, and hence, our solution can estimate both depth map and parameters by information propagation on the dynamic Bayesian network spreading to the image resolution direction. In this scheme, a strategy with low computational cost is quite important, since simultaneous complete belief propagation (BP) is a complicated problem. Therefore, we adopt suitable approximations, i.e. the Laplace approximation is applied to the posterior probabilities of the parameters and the saddle point approximation is introduced to the posterior of the depth, and additionally, the Supplemented MAP-EM algorithm (Meng & Rubin, 1991) is applied as a stable and effective processing. The remarkable feature of the Supplemented MAP-EM algorithm compared with the simple MAP-EM algorithm is that by the Supplemented MAP-EM algorithm the asymptotic variance-covariance of the parameters can be estimated using only the code for the MAP-EM algorithm itself and the code for standard matrix operation instead of the numerical differentials. The asymptotic variance-covariance matrix is necessary for BP, and then, the application of the Supplemented MAP-EM algorithm is a strong assertion in this study. This framework can be used for a lot of fields in which the usual Kalman filter can be applied effectively.

4. Computation Principle based on Multi-Scale Processing

4.1 Image decomposition and probabilistic models

Firstly, we explain decomposition of observed images into multi-scale images. In order to simply decompose those into multiple images having different resolutions, it is reasonable to vary the cut-off frequency of the spatial low-pass filter. However, since we assume for the following processing that the observation errors contained in the temporal differentials of the image intensity have statistically independency among different resolutions, we adopt the decomposition using a spatial band-pass filter. In the following, index $l = 1, 2, \dots, L$ represents the difference of the frequency band of the image, and $l = 1$ indicates the lowest resolution image, called the lowest image hereafter. The resolution of the depth map for each resolution image is defined according to its frequency band which corresponds to the spatial wavelength. Namely, for the image indexed by l , depth is assumed to be constant in a local region which size is N_l , and size N_l is determined along with a rule that $N_{l_1} < N_{l_2}$ if

$l_1 > l_2$. Although a pyramid structure of multi-scale images, for example the Wavelet transform, is suitable for effective treatment of the information, all information observed at all pixels is used without performing down sampling.

Next, we define the probabilistic models. As mentioned in Sec. 2.2, translational vector \mathbf{u} is described as $\mathbf{u} = [u_x, u_y, (1-u_x^2-u_y^2)^{1/2}]^T$, hence the scale of d should be fixed by this normalization of \mathbf{u} . As mentioned above, d is a constant in the local region for each resolution image, and the constants for all the local regions are independent. Among the different resolutions, d is supposed to be a conditionally independent unknown variable. This is based on the definition

$$d^{(l+1)} = I^{(l)}[d^{(l)}] + n_0^{(l)}, \quad (7)$$

where $I^{(l)}$ indicates a linear interpolation operator from l to $l+1$. In this equation, $n_0^{(l)}$ is a Gaussian random variable with zero mean and variance σ_0^2 common to all local regions and all resolutions, which means a perturbation from the interpolation value and it is statistically independent of other random variables. The probability density function of $d^{(l+1)}$ is given as

$$p(d^{(l+1)} | \sigma_0^2) = \frac{1}{\sqrt{2\pi\sigma_d^{2(l+1)}}} \exp\left[-\frac{(d^{(l+1)} - m_d^{(l+1)})^2}{2\sigma_d^{2(l+1)}}\right], \quad (8)$$

where the mean $m_d^{(l+1)}$ and the variance $\sigma_d^{2(l+1)}$ can be formulated recursively as follows:

$$m_d^{(l+1)} = I^{(l)}[m_d^{(l)}], \quad (9)$$

$$\sigma_d^{2(l+1)} = I^{(l)2}[\sigma_d^{2(l)}] + \sigma_0^2. \quad (10)$$

Equation 10 indicates an approximated representation in which the covariance terms of $d^{(l)}$ are neglected and hence only the variance terms are considered. Moreover, $I^{(l)2}$ is also the linear interpolation operator, the weight coefficients of which correspond to the power of each of the corresponding coefficient of $I^{(l)}$.

Subsequently, we define a probabilistic model for an observation f_t . An observation equation containing an observation error is given based on Eq. 6 as follows:

$$f_t^{(l)} = -f^{r(l)} - f^{u(l)} d^{(l)} + n_1^{(l)}. \quad (11)$$

There have been several discussions with respect to a noise model of Eq. 11 (Nestares et al., 2000; Weiss & Fleet, 2001). For simplicity, the observation error $n_1^{(l)}$ is a Gaussian random variable with zero mean and variance σ_1^2 common to all local regions and all resolutions

and it is statistically independent of other random variables. From Eq. 11, $f_t^{(l)}$ is also a Gaussian random variable and its conditional probability density can be formulated as

$$p(f_t^{(l)} | d^{(l)}, \sigma_1^2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp \left[-\frac{(f_t^{(l)} + f^{r(l)} + f^{u(l)} d^{(l)})^2}{2\sigma_1^2} \right]. \quad (12)$$

The probabilistic model defined above can be represented as a graphical model, i.e. the Bayesian network shown in Fig. 2. In this network, the parameters $\Theta \equiv \{u_x, u_y, r, \sigma_0^2, \sigma_1^2\}$, which are also shown in Fig. 2, are regarded as probabilistic variables estimated through BP, which is described in the following section. In this figure, the parameters are shown not to be common to the layers formally, but they are assumed to be constants with respect to the layers in this study.

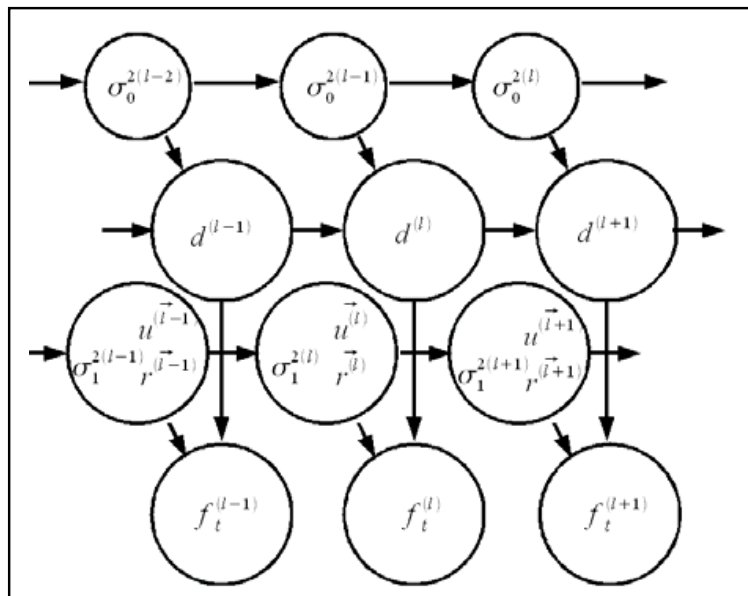


Fig. 2. Graphical model used in this chapter

4.2 Hierarchical estimation of depth

Based on the probabilistic models defined in Sec. 4.1, we are going to find a minimum variance estimator $\hat{d}^{(l+1)}$, when $\{f_t^{(l+1)}\}, \{f_t^{(l)}\}, \dots, \{f_t^{(1)}\}$ are observed, where the symbol $\{a\}$ represents a set of arbitrary values a defined at all the pixels. Such $\hat{d}^{(l+1)}$ corresponds to the mean of a posterior probability of $d^{(l+1)}$ after obtaining all observations. This posterior probability is introduced as follows.

Let $\{f_t^{(l)}\}_N$ be a set of $f_t^{(l)}$ in a local region N in an image where $d^{(l)}$ is assumed to be constant, and the next holds

$$p(d^{(l+1)} | \{f_t^{(l+1)}\}, \{f_t^{(l)}\}, \dots) = p(d^{(l+1)} | \{f_t^{(l+1)}\}_N, \{f_t^{(l)}\}, \dots) \quad (13)$$

This can be rewritten using the Bayes formula as follows:

$$p(d^{(l+1)} | \{f_t^{(l+1)}\}_N, \{f_t^{(l)}\}, \dots) = \frac{p(\{f_t^{(l+1)}\}_N | d^{(l+1)}) p(d^{(l+1)} | \{f_t^{(l)}\}, \dots)}{p(\{f_t^{(l+1)}\}_N | \{f_t^{(l)}\}, \dots)}. \quad (14)$$

The numerator of the right hand side of Eq. 14 can be concretely shown. Firstly, the second term of this numerator can be written using $d^{(l)}$ which resolution is one-step lower than $d^{(l+1)}$

$$p(d^{(l+1)} | \{f_t^{(l)}\}, \dots) = \int p(d^{(l+1)} | \{d^{(l)}\}_I) p(\{d^{(l)}\}_I | \{f_t^{(l)}\}, \dots) d\{d^{(l)}\}_I, \quad (15)$$

$$p(\{d^{(l)}\}_I | \{f_t^{(l)}\}, \dots) = \prod_I p(d^{(l)} | \{f_t^{(l)}\}_N, \dots), \quad (16)$$

where $\{\cdot\}_I$ indicates a set of the elements used for the linear interpolation. From Eqs. 7 and 11, Eq. 15 also represents the Gaussian distribution, and using the mean $\tilde{d}^{(l)}$ and the variance $\tilde{\sigma}_d^{2(l)}$ of the posterior distribution $p(d^{(l)} | \{f_t^{(l)}\}, \dots)$ for resolution l , Eq. 15 can be represented as Eq. 8, i.e.,

$$p(d^{(l+1)} | \{f_t^{(l)}\}, \dots) = \frac{1}{\sqrt{2\pi(I^{(l)2}[\tilde{\sigma}_d^{2(l)}] + \sigma_0^2)}} \exp\left[-\frac{(d^{(l+1)} - I^{(l)}[\tilde{d}^{(l)}])^2}{2(I^{(l)2}[\tilde{\sigma}_d^{2(l)}] + \sigma_0^2)}\right]. \quad (17)$$

Therefore, the numerator of Eq. 14 can be formulated as follows:

$$p(\{f_t^{(l+1)}\}_N | d^{(l+1)}) p(d^{(l+1)} | \{f_t^{(l)}\}, \dots) = \left[\prod_N p(f_t^{(l+1)} | d^{(l+1)}) \right] p(d^{(l+1)} | \{f_t^{(l)}\}, \dots) = \frac{\exp(-H)}{Z}, \quad (18)$$

$$Z = \left(\sqrt{2\pi\sigma_1^2}\right)^N \sqrt{2\pi(I^{(l)2}[\tilde{\sigma}_d^{2(l)}] + \sigma_0^2)}, \quad (19)$$

$$H = \frac{\sum_N \left(f_t^{(l+1)} + f_r^{(l+1)} + f_u^{(l+1)} d^{(l+1)}\right)^2}{2\sigma_1^2} + \frac{(d^{(l+1)} - I^{(l)}[\tilde{d}^{(l)}])^2}{2(I^{(l)2}[\tilde{\sigma}_d^{2(l)}] + \sigma_0^2)} = \frac{(d^{(l+1)} - \tilde{d}^{(l+1)})^2}{2\tilde{\sigma}_d^{2(l+1)}} + \text{const.}, \quad (20)$$

$$\tilde{d}^{(l+1)} = \tilde{\sigma}_d^{2(l+1)} \left(\frac{\mathbf{I}^{(l)} [\tilde{d}^{(l)}]}{\mathbf{I}^{(l)2} [\tilde{\sigma}_d^{2(l)}] + \sigma_0^2} - \frac{\sum_N (f_t^{(l+1)} + f_r^{(l+1)}) f_u^{(l+1)}}{\sigma_1^2} \right), \quad (21)$$

$$\tilde{\sigma}_d^{2(l+1)} = \left(\frac{1}{\mathbf{I}^{(l)2} [\tilde{\sigma}_d^{2(l)}] + \sigma_0^2} + \frac{\sum_N f_u^{(l+1)2}}{\sigma_1^2} \right)^{-1}, \quad (22)$$

where N is the number of pixels in the region N . The estimator of $d^{(l+1)}$ minimizing H formulated by Eq. 20 is the MAP (Maximum A Posteriori) estimator $d_{MAP}^{(l+1)}$ and it coincides with the mean of the posterior probability in Eq. 14. Hence, $d_{MAP}^{(l+1)} = \tilde{d}^{(l+1)}$ holds. For the Gaussian distribution, the MAP estimator is also the minimum variance estimator. It can be known that $\tilde{d}^{(l+1)}$ is computed recursively using Eq. 21, and this procedure is an essential operation of the Kalman filter. Equations 7 and 11 correspond to the state equation and the observation equation, respectively.

The Kalman filter is usually used to estimate effectively a marginal posterior probability of inner state at the present time using observations already measured. Since we aim to estimate $\{d^{(L)}\}$, and $\{d^{(l)}\} (l < L)$ is not explicitly needed, in the proposed scheme, we can determine $\{d^{(L)}\}$ in a way similar to the Kalman filter strategy, i.e. we use the marginal posterior probability $p(\{d^{(L)}\} | \{f_t^{(L)}\}, \{f_t^{(L-1)}\}, \dots)$ instead of the simultaneous posterior probability $p(\{d^{(L)}\}, \{d^{(L-1)}\}, \dots | \{f_t^{(L)}\}, \{f_t^{(L-1)}\}, \dots)$ to determine $\{d^{(L)}\}$. This estimator is often called the MPM (Marginal Posterior Mode) estimator. Another essential reason for using the Kalman filter strategy is the fact that the alias problem should be avoided. When $f_t^{(l)}$ is measured as a difference using two successive frames, if the amplitude of the 2-D motion is larger than the spatial wavelength corresponding to the spatial frequency band of $f^{(l)}$, the undesirable aliasing occurs. Therefore, we define $f_t^{(l)}$ in which aliasing can not be seen using the estimated optical flow $\hat{\mathbf{v}}^{(l-1)} = [\hat{v}_x^{(l-1)}, \hat{v}_y^{(l-1)}]^T$ calculated indirectly from the depth $\tilde{d}^{(l-1)}$ obtained for the one-step low resolution images $f^{(l-1)}$ as follows:

$$f_t^{(l)} = -f_x^{(l)} \mathbf{I}^{(l-1)} [\hat{v}_x^{(l-1)}] - f_y^{(l)} \mathbf{I}^{(l-1)} [\hat{v}_y^{(l-1)}] + \frac{\partial}{\partial t} W[f^{(l)}, \mathbf{I}^{(l-1)} [\hat{\mathbf{v}}^{(l-1)}]], \quad (23)$$

where a warp operation is defined as

$$W[f, \mathbf{v}](\mathbf{x}, t + \delta t) \equiv f(\mathbf{x} - \mathbf{v} \delta t, t + \delta t), \quad (24)$$

and partial differential $\partial / \partial t$ is done as a finite difference. From the above definitions, the optical flow estimations at the lower resolution step, which estimations have little risk of aliasing, should be used successively, so as to avoid aliasing and detect stable $f_t^{(l)}$. For such

purpose, the sequential procedure from low resolution step to high resolution step is required.

4.3 Computation flow including parameter determination

For the successive estimation described in the above section, there are some parameters which should be known in advance. It is usual that parameters are treated as definite variables and are determined as a maximum likelihood (ML) estimator. On the other hand, it is known that the MAP estimator obtained by considering a parameter as a probabilistic variable formally having a uniform distribution coincides with a ML estimator. Hence, in this study, we suppose that the parameters are probabilistic variables as well as depth.

Since we assume that σ_0^2 and σ_1^2 are common to all resolution steps, $\Theta = \{u_x, u_y, \mathbf{r}, \sigma_0^2, \sigma_1^2\}$ has to be determined with no dependence of each resolution processing. The information of $\{d^{(l)}\}$ is propagated from low resolution to high resolution, therefore we estimate Θ by the same scheme and adopt the Bayesian inference formally supposing a prior of Θ . The posterior probability density of Θ can be decomposed in the following way

$$p(\Theta | \{f_t^{(L)}\}, \dots, \{f_t^{(1)}\}) \quad (25)$$

$$= \frac{p(\{f_t^{(L)}\} | \{f_t^{(L-1)}\}, \dots, \{f_t^{(1)}\}, \Theta) p(\{f_t^{(L-1)}\} | \{f_t^{(L-2)}\}, \dots, \{f_t^{(1)}\}, \Theta) \dots p(\{f_t^{(1)}\} | \Theta) p(\Theta)}{p(\{f_t^{(L)}\} | \{f_t^{(L-1)}\}, \dots, \{f_t^{(1)}\}) p(\{f_t^{(L-1)}\} | \{f_t^{(L-2)}\}, \dots, \{f_t^{(1)}\}) \dots p(\{f_t^{(1)}\})}.$$

We assume that the last term at the right hand side of Eq. 25, i.e. $p(\Theta | \{f_t^{(1)}\})$, is explicitly known. After observing $\{f_t^{(2)}\}$ from the next higher resolution images, we can compute

$$\frac{p(\{f_t^{(2)}\} | \{f_t^{(1)}\}, \Theta)}{p(\{f_t^{(2)}\} | \{f_t^{(1)}\})} p(\Theta | \{f_t^{(1)}\}) = \frac{p(\{f_t^{(2)}\}, \Theta | \{f_t^{(1)}\})}{p(\{f_t^{(2)}\} | \{f_t^{(1)}\})} = p(\Theta | \{f_t^{(2)}\}, \{f_t^{(1)}\}) \quad (26)$$

In this equation, the term $p(\{f_t^{(2)}\} | \{f_t^{(1)}\}, \Theta)$ can be derived using Eqs. 12 and 17 as follows:

$$p(\{f_t^{(2)}\} | \{f_t^{(1)}\}, \Theta) = \int p(\{f_t^{(2)}\} | \{d^{(2)}\}, \Theta) p(\{d^{(2)}\} | \{f_t^{(1)}\}, \Theta) d\{d^{(2)}\} \quad (27)$$

It should be noticed that Θ is omitted in Eqs. 12 and 17 and in these equations Θ means a true value, but in Eq. 27 Θ is considered as a variable. By propagating the computation of Eq. 26 successively from low resolution images to high resolution images, we can finally obtain the left hand side of Eq. 25. This procedure also coincides with Kalman filtering for a state variable having no dynamic transition. By assuming a prior having a large entropy for $p(\Theta)$, for example a uniform distribution, the Bayesian estimator Θ_{MAP} approximately equals to the ML estimator.

We again refer the successive estimation of $\{d^{(l)}\}$, and also in Eqs. 14, 21 and 22 the true value of Θ is necessary to be known. Correctly, Eq. 14 should be written as $p(d^{(l+1)}|\{f_t^{(l+1)}\}_N, \dots, \{f_t^{(1)}\}, \Theta)$. Hence, in order to solve a problem, in which Θ is unknown as well as $\{d^{(L)}\}$, the marginalization with respect to Θ is required to propagate the information of depth as follows:

$$p(d^{(l+1)}|\{f_t^{(l+1)}\}_N, \{f_t^{(l)}\}, \dots, \{f_t^{(1)}\}) = \int p(d^{(l+1)}|\{f_t^{(l+1)}\}_N, \{f_t^{(l)}\}, \dots, \{f_t^{(1)}\}, \Theta) p(\Theta|\{f_t^{(l+1)}\}, \dots, \{f_t^{(1)}\}) d\Theta. \quad (28)$$

From the above discussion, the flow of procedures at each resolution step is summarized in the following five steps. We assume that at the l th layer the posteriors of Θ and $\{d^{(l)}\}$ are already derived as $p(\Theta|\{f_t^{(l)}\}, \dots, \{f_t^{(1)}\})$ and $p(d^{(l)}|\{f_t^{(l)}\}_N, \dots, \{f_t^{(1)}\})$, respectively.

- (i) Using Eq. 17, compute the predictive posterior probability $p(d^{(l+1)}|\{f_t^{(l)}\}, \dots, \{f_t^{(1)}\}, \Theta)$ from $p(d^{(l)}|\{f_t^{(l)}\}_N, \dots, \{f_t^{(1)}\})$.
- (ii) By generalizing Eqs. 26 and 27 for $(l+1)$ th layer, compute the posteriors of the parameters $p(\Theta|\{f_t^{(l+1)}\}, \dots, \{f_t^{(1)}\})$ from $p(\Theta|\{f_t^{(l)}\}, \dots, \{f_t^{(1)}\})$ at the previous layer, $p(d^{(l+1)}|\{f_t^{(l)}\}, \dots, \{f_t^{(1)}\}, \Theta)$ obtained at (i) and $p(f_t^{(l+1)}|d^{(l+1)}, \Theta)$ defined in Eq. 12.
- (iii) Using $p(d^{(l+1)}|\{f_t^{(l)}\}, \dots, \{f_t^{(1)}\}, \Theta)$ and $p(f_t^{(l+1)}|d^{(l+1)}, \Theta)$ alike in (ii), compute the posterior of depth conditioned on the parameters $p(d^{(l+1)}|\{f_t^{(l+1)}\}_N, \dots, \{f_t^{(1)}\}, \Theta)$.
- (iv) Using Eq. 28, compute the marginal posterior of depth $p(d^{(l+1)}|\{f_t^{(l+1)}\}_N, \dots, \{f_t^{(1)}\})$ from $p(d^{(l+1)}|\{f_t^{(l+1)}\}_N, \dots, \{f_t^{(1)}\}, \Theta)$ obtained in (iii) and $p(\Theta|\{f_t^{(l+1)}\}, \dots, \{f_t^{(1)}\})$ obtained in (ii).
- (v) The above procedures should be repeated for the next resolution layer.

There is another scheme to determine the parameters besides the above one. By considering both $\{d^{(l)}\}$ and Θ as state variables, their successive updating can be performed. Based on the extended Kalman filter theory, this can be formulated by linearizing locally the observation equation and the state equation with respect to Θ . However, the obtained estimator is the Bayesian estimator based on the simultaneous posterior $p(\{d^{(l)}\}, \Theta|\{f_t^{(l)}\}, \dots, \{f_t^{(1)}\})$. On the other hand, our scheme can obtain the estimators based on both marginal posteriors $p(\{d^{(l)}\}|\{f_t^{(l)}\}, \dots, \{f_t^{(1)}\})$ and $p(\Theta|\{f_t^{(l)}\}, \dots, \{f_t^{(1)}\})$. Since $\{d^{(l)}\}$ and Θ have no special meaning as a pair, the estimation scheme in our study is adequate.

5. Computing Algorithm

5.1 Application of EM algorithm for approximation

In order to approximately execute the BP procedure mentioned in the above section, we can use a framework with the EM algorithm (Dempster et al., 1977). The EM algorithm is an effective scheme if applied to the problem which is easy to be solved if some unknown variables, often called the "hidden variables", are observed in addition to actual observations. Such hidden variables and observations considered together are called the "complete data," and the observations themselves are called the "incomplete data." By the EM algorithm, the posterior probabilities of the hidden variables and the ML estimators of the parameters can be obtained through iterative procedures. Additionally, by the MAP-EM algorithm extended version of EM algorithm, the MAP estimation of the parameters can be performed based on the posterior probabilities of the parameters marginalized with respect to the hidden variables. Although the details are omitted, the following two steps are executed at the each iteration:

[E-step]: The posterior probabilities of the hidden variables are derived using the parameters values estimated at the previous iteration. Using these probabilities the evaluation function, called Q function, needed for the parameters updating is introduced.

[M-step]: The parameters values are updated by maximizing the Q function introduced in the E-step.

In this study, the MAP-EM algorithm is applied for each resolution l , and hence, the MAP estimator Θ_{MAP} using the prior $p(\Theta | \{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\})$ and the MAP estimator $\{d_{MAP}^{(l)}\}$ based on its posterior conditioned by Θ_{MAP} are determined. These estimators are not completely the ones given in Eqs. 25 and 28. However, the application of the MAP-EM algorithm with supplemented function described in the following subsections in details enables an efficient computation algorithm for our problem, and the estimators derived by this algorithm can be assumed to be appropriate approximations.

5.2 MAP-EM algorithm for determining depth and motion

In this section, we introduce the explicit formulations for the resolution l , i.e. the layer l , to obtain Θ_{MAP} and $\{d_{MAP}^{(l)}\}$ using the images which resolutions are not exceeding l . At the E-step, the following Q function with respect to Θ is constructed

$$Q(\Theta; \hat{\Theta}) = E \left[\ln p(\{f_t^{(l)}\}, \{d^{(l)}\} | \{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}, \Theta) + \ln p(\Theta | \{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}) \mid \{f_t^{(l)}\}, \dots, \{f_t^{(1)}\}, \hat{\Theta} \right], \quad (29)$$

where $E \left[\cdot \mid \{f_t^{(l)}\}, \dots, \{f_t^{(1)}\}, \hat{\Theta} \right]$ denotes the conditional expectation using $p(\{d^{(l)}\} | \{f_t^{(l)}\}, \dots, \{f_t^{(1)}\}, \hat{\Theta})$, and in the following formulations including Eq. 29, the symbol $\hat{\cdot}$ indicates the estimate or variable depending on the estimate derived in this iteration. In Eq. 29, the simultaneous probability of $\{f_t^{(l)}\}$ and $\{d^{(l)}\}$ is

$$\begin{aligned}
 p(\{f_t^{(l)}\}, \{d^{(l)}\} | \{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}, \Theta) &= p(\{f_t^{(l)}\} | \{d^{(l)}\}, \Theta) p(\{d^{(l)}\} | \{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}, \Theta) \\
 &= \prod_{i=1}^{M^{(l)}} \left[\prod_{N_i^{(l)}} p(f_t^{(l)} | d_i^{(l)}, \Theta) \right] \prod_{i=1}^{M^{(l)}} p(d_i^{(l)} | \{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}, \Theta)
 \end{aligned} \quad (30)$$

In Eq. 30, we suppose that the number of pixels in the local region $N_i^{(l)}$ is constant with respect to the local region index i and it takes value $N^{(l)}$. Additionally, the number of the local regions in an image is given by $M^{(l)}$. The function which expectation is computed in Eq. 29 is concretely written using Eqs. 12 and 17 as follows:

$$\begin{aligned}
 & -\frac{N^{(l)}M^{(l)}}{2} \ln \sigma_1^2 - \frac{1}{2} \sum_{i=1}^{M^{(l)}} \ln \left(\mathbf{I}^{(l-1)^2} \left[\tilde{\sigma}_d^{2(l-1)} \right]_i + \sigma_0^2 \right) - \frac{1}{2\sigma_1^2} \sum_{i=1}^{M^{(l)}} \sum_{N_i^{(l)}} \left(f_t^{(l)} + f^{r(l)} + f^{u(l)} d_i^{(l)} \right)^2 \\
 & - \frac{1}{2} \sum_{i=1}^{M^{(l)}} \frac{\left(d_i^{(l)} - \mathbf{I}^{(l-1)} \left[\tilde{d}^{(l-1)} \right]_i \right)^2}{\left(\mathbf{I}^{(l-1)^2} \left[\tilde{\sigma}_d^{2(l-1)} \right]_i + \sigma_0^2 \right)} - \frac{1}{2} (\Theta - \tilde{\Theta})^T \tilde{\mathbf{V}}^{-1} (\Theta - \tilde{\Theta}) + \text{Const.},
 \end{aligned} \quad (31)$$

where the Laplace approximation is applied to $p(\Theta | \{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\})$, i.e. it is approximated by a Gaussian distribution, since it has a complex form. To simplify the computations, we assume that there is no correlation between each two of $\mathbf{m} \equiv \{u_x, u_y, \mathbf{r}\}$, σ_0^2 and σ_1^2 . As parameters of the Laplace approximation, the mean $\tilde{\Theta}$ and the covariance $\tilde{\mathbf{V}}$ of Θ at layer $l-1$ are used. Their estimation method is described in the next section.

On the other hand, $p(d_i^{(l)} | \{f_t^{(l)}\}_{N_i}, \dots, \{f_t^{(1)}\}, \hat{\Theta})$ used for the expectation in Eq. 29 has a Gaussian distribution, and its mean and variance are represented using $\hat{\Theta}$ and Eqs. 21 and 22

$$\begin{aligned}
 \hat{d}_i^{(l)} &= \hat{\sigma}_{d_i}^{2(l)} \left(\frac{\mathbf{I}^{(l-1)} \left[\tilde{d}^{(l-1)} \right]_i}{\mathbf{I}^{(l-1)^2} \left[\tilde{\sigma}_d^{2(l-1)} \right]_i + \hat{\sigma}_0^2} - \frac{\sum_{N_i^{(l)}} (f_t^{(l)} + \hat{f}^{r(l)}) \hat{f}^{u(l)}}{\hat{\sigma}_1^2} \right), \\
 \hat{\sigma}_{d_i}^{2(l)} &= \left(\frac{1}{\mathbf{I}^{(l-1)^2} \left[\tilde{\sigma}_d^{2(l-1)} \right]_i + \hat{\sigma}_0^2} + \frac{\sum_{N_i^{(l)}} \hat{f}^{u(l)^2}}{\hat{\sigma}_1^2} \right)^{-1}.
 \end{aligned} \quad (32) \quad (33)$$

By taking the expectation of Eq. 31 with respect to $p(d_i^{(l)} | \{f_t^{(l)}\}_{N_i}, \dots, \{f_t^{(1)}\}, \hat{\Theta})$ using Eqs. 32 and 33, $Q(\Theta; \hat{\Theta})$ in Eq. 29 can be concretely derived. At the M-step, $Q(\Theta; \hat{\Theta})$ should be maximized with respect to Θ in order to update Θ . Therefore, to update Θ we can minimize the following function, which is obtained by multiplying $Q(\Theta; \hat{\Theta})$ by -2 and neglecting the constant value

$$J(\Theta) = N^{(l)} M^{(l)} \ln \sigma_1^2 + \sum_{i=1}^{M^{(l)}} \ln \left(\mathbf{I}^{(l-1)^2} \left[\tilde{\sigma}_d^{2(l-1)} \right]_i + \sigma_0^2 \right) + J_f(\mathbf{m}, \sigma_1^2) + J_d(\sigma_0^2) + J_p(\Theta), \quad (34)$$

$$J_f(\mathbf{m}, \sigma_1^2) \quad (35)$$

$$= \frac{1}{\sigma_1^2} \left\{ \sum_{i=1}^{M^{(l)}} \sum_{N_i^{(l)}} \left(f_t^{(l)} + f^{r(l)} \right)^2 + 2 \sum_{i=1}^{M^{(l)}} \hat{d}_i^{(l)} \sum_{N_i^{(l)}} f^{u(l)} \left(f_t^{(l)} + f^{r(l)} \right) + \sum_{i=1}^{M^{(l)}} \left(\hat{d}_i^{(l)^2} + \hat{\sigma}_{d_i}^{2(l)} \right) \sum_{N_i^{(l)}} f^{u(l)^2} \right\},$$

$$J_d(\sigma_0^2) = \sum_{i=1}^{M^{(l)}} \frac{\hat{d}_i^{(l)^2} + \hat{\sigma}_{d_i}^{2(l)} - 2 \hat{d}_i^{(l)} \mathbf{I}^{(l-1)} \left[\tilde{d}^{(l-1)} \right]_i + \left(\mathbf{I}^{(l-1)} \left[\tilde{d}^{(l-1)} \right]_i \right)^2}{\mathbf{I}^{(l-1)^2} \left[\tilde{\sigma}_d^{2(l-1)} \right]_i + \sigma_0^2}, \quad (36)$$

$$J_p(\Theta) = (\mathbf{m} - \tilde{\mathbf{m}})^T \tilde{\mathbf{V}}_m^{-1} (\mathbf{m} - \tilde{\mathbf{m}}) + \frac{(\sigma_0^2 - \tilde{\sigma}_0^2)^2}{\tilde{\sigma}_{\sigma 0}^2} + \frac{(\sigma_1^2 - \tilde{\sigma}_1^2)^2}{\tilde{\sigma}_{\sigma 1}^2}. \quad (37)$$

Minimization of $J(\Theta)$ in Eq. 34 cannot be done analytically; therefore, numerical search has to be performed. In general, the value completely minimizing an objective function is difficult to be found. In such case, we can use the generalized MAP-EM algorithm in which, at the M-step, parameter updating is done so as to enlarge the value of the Q function more than that for the parameters values obtained at the previous iteration. By the generalized MAP-EM algorithm, the computational cost for each M-step often decreases, although the number of iterations may increase.

After convergence of the above two steps for each layer l , if the initial values of Θ for the iteration are suitable, we can obtain Θ_{MAP} , which maximizes $p(\Theta | \{f_t^{(l)}\}, \dots, \{f_t^{(1)}\})$ and coincides with the mean $\tilde{\Theta}$ of this probability because of the Laplace approximation, and $d_{MAP}^{(l)}$, which maximizes $p(d^{(l)} | \{f_t^{(l)}\}_N, \dots, \{f_t^{(1)}\}, \tilde{\Theta})$ and also coincides with the mean $\tilde{d}^{(l)}$. The probability which should be actually evaluated corresponds to Eq. 28, and the integration for this marginalization requires a numerical computation or a random sampling technique. Hence, in this study, we justify the solution of the above MAP-EM algorithm by applying the saddle point approximation to $p(\Theta | \{f_t^{(l)}\}, \dots, \{f_t^{(1)}\})$, as follows:

$$p(d^{(l)} | \{f_t^{(l)}\}_N, \dots, \{f_t^{(1)}\}) \cong \int p(d^{(l)} | \{f_t^{(l)}\}_N, \dots, \{f_t^{(1)}\}, \Theta) \delta(\Theta - \hat{\Theta}) d\Theta \quad (38)$$

$$= p(d^{(l)} | \{f_t^{(l)}\}_N, \dots, \{f_t^{(1)}\}, \hat{\Theta})$$

5.3 Parameter variance estimation using Supplemented EM technique

In order to evaluate Eq. 34, i.e. Eq. 37, for each layer l , the variance-covariance matrix of Θ at the previous layer $l-1$ is required. However, this matrix cannot be estimated directly by the MAP-EM algorithm. Usually used naive approximate method is evaluating

$\left(-\partial^2 \ln p(\Theta | \{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}) / \partial \Theta^2\right)^{-1}$ numerically at Θ_{MAP} of layer $l-1$. In this study, we aim to compute the variance-covariance matrix by a stable and efficient scheme, and an application of the Supplemented EM (SEM) algorithm (Meng & Rubin, 1991) is examined and proposed. Using the SEM algorithm, the asymptotic variance-covariance of the parameters can be estimated using only the code for the EM algorithm and the code for computing the complete data asymptotic variance-covariance matrix.

Since, in this study, the priors of the parameters should be considered at each layer l , the usage of the MAP-EM algorithm is appropriate. Hence, the S-MAP-EM algorithm, which is the MAP-EM algorithm with the supplemented procedures described below in detail, is actually used instead of the SEM algorithm. Let $\mathbf{V}_{\hat{\Theta}}$ denote the “observed” asymptotic variance-covariance matrix, which is evaluated for the converging value $\hat{\Theta}$ without expectation operation, and it is used as an estimate. The important equation for the S-MAP-EM algorithm is

$$\mathbf{V}_{\hat{\Theta}} = \mathbf{V}_{c\hat{\Theta}} + \delta \mathbf{V}, \quad (39)$$

$$\mathbf{V}_{c\hat{\Theta}} = \left\{ -\frac{\partial^2}{\partial \Theta \partial \Theta^T} \left[\ln p(\{f_t^{(l)}\}, \{d^{(l)}\} | \{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}, \Theta) + \ln p(\Theta | \{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}) \right]_{\Theta=\hat{\Theta}} \right\}^{-1}, \quad (40)$$

$$\delta \mathbf{V} = \{\mathbf{I} - \mathbf{DM}_{\hat{\Theta}}\}^{-1} \mathbf{DM}_{\hat{\Theta}} \mathbf{V}_{c\hat{\Theta}}. \quad (41)$$

In Eq. 41, \mathbf{I} indicates a unit matrix, and \mathbf{DM} is the Jacobian matrix for the mapping $M: \Theta^k \rightarrow \Theta^{k+1}$ implicitly defined by the MAP-EM algorithm, where k indicates the iteration number. In the same way as a wide class of problems where probability of the complete-data belongs to the exponential family, $\mathbf{V}_{c\hat{\Theta}}$ can be derived analytically in this study. On the other hand, $\mathbf{DM}_{\hat{\Theta}}$ has to be estimated by a stable computation.

If Θ^k converges to some value $\hat{\Theta}$ and $M(\Theta)$ is continuous, $\hat{\Theta}$ is a fixed point of the iteration and satisfies $\hat{\Theta} = M(\hat{\Theta})$. By a Taylor series expansion of $\Theta^{k+1} = M(\Theta^k)$ at $\hat{\Theta}$, we can have the first approximation

$$\Theta^{k+1} - \hat{\Theta} \approx \mathbf{DM}_{\hat{\Theta}} (\Theta^k - \hat{\Theta}) \quad (42)$$

From this equation, the $d \times d$ matrix $\mathbf{DM}_{\hat{\Theta}}$ is often referred to as the rate of convergence, where d is the degree of freedom of Θ and in this study $d=7$. It is convenient to estimate $\mathbf{DM}_{\hat{\Theta}}$ by numerical evaluation using an iteration procedure, for example the following formulation

$$\lim_{k \rightarrow \infty} \frac{\Theta_i^{k+1} - \Theta_i^k}{\Theta_i^k - \Theta_i^{k-1}}, \quad i = 1, \dots, d. \quad (43)$$

However, this provides only a few eigenvalues of $\mathbf{DM}_{\hat{\Theta}}$, and the matrix itself cannot be computed. On the other hand, Meng and Rubin (1991) explained that each element of $\mathbf{DM}_{\hat{\Theta}}$ is the component-wise rate of convergence of a “forced EM” for the S-MAP-EM algorithm. Let r_{ij} be the (i, j) th element of $\mathbf{DM}_{\hat{\Theta}}$, and we define

$$\Theta_{(j)}^k \equiv [\hat{\Theta}_1, \dots, \hat{\Theta}_{j-1}, \Theta_j^k, \hat{\Theta}_{j+1}, \dots, \hat{\Theta}_d]^T, \quad (44)$$

where Θ_j^k is the value of Θ_j on the k th iteration of the MAP-EM algorithm and $\hat{\Theta}_i$ indicates the i th component of the converging value $\hat{\Theta}$. By the definition of r_{ij} , the following equation holds

$$\begin{aligned} r_{ij} &= \frac{\partial M_i(\hat{\Theta})}{\partial \Theta_j} \\ &= \lim_{\Theta_j \rightarrow \hat{\Theta}_j} \frac{M_i(\hat{\Theta}_1, \dots, \hat{\Theta}_{j-1}, \Theta_j, \hat{\Theta}_{j+1}, \dots, \hat{\Theta}_d) - M_i(\hat{\Theta})}{\Theta_j - \hat{\Theta}_j} \\ &= \lim_{k \rightarrow \infty} \frac{M_i(\Theta_{(j)}^k) - \hat{\Theta}_i}{\Theta_j^k - \hat{\Theta}_j} \\ &\equiv \lim_{k \rightarrow \infty} r_{ij}^k. \end{aligned} \quad (45)$$

From this, for example, the following procedure can be introduced to obtain $\mathbf{V}_{c\hat{\Theta}}$.

S-MAP-EM algorithm

- (i) Run the MAP-EM algorithm in order to obtain $\hat{\Theta}$, and save all the values $\{\Theta^k\}$ at each iteration.
- (ii) For each Θ^k , compute $\Theta_{(j)}^k$ using Eq. 44, and run one iteration of the MAP-EM algorithm using $\Theta_{(j)}^k$ as a current estimate to obtain $M_i(\Theta_{(j)}^k)$, and subsequently the ratio r_{ij}^k in Eq. 45 for each i ($i = 1, \dots, d$). This is done for each j ($j = 1, \dots, d$).
- (iii) Check the convergence of r_{ij}^k using $|r_{ij}^k - r_{ij}^{k+1}| < \xi$ with a certain threshold value ξ , and determine r_{ij} ($i, j = 1, \dots, d$) and hence, $\mathbf{DM}_{\hat{\Theta}}$.
- (iv) Compute $\mathbf{V}_{c\hat{\Theta}}$ by Eq. 40, and then using Eqs. 39 and 41, evaluate the observed asymptotic variance-covariance matrix $\mathbf{V}_{\hat{\Theta}}$.

At step (i) in the above procedure, we assume that the updating sequence of the parameters by the MAP-EM algorithm is stored to save computational time. However, to save extra storage, at step (ii) one-iteration of the MAP-EM algorithm to obtain Θ^k from Θ^{k-1} is done firstly instead of saving all the values of Θ^k . Using this Θ^k , one-iteration of the S-MAP-EM at step (ii) is realized.

Note that δV is a symmetric matrix, and if it seems to be quite asymmetric, there has been a programming error in either MAP-EM or S-MAP-EM, or convergences of both have not been sufficient.

6. Implementation

6.1 Image decomposition into multi-scale images

Ideally, we have to decompose images into multi-scale images using spatio-temporal filtering. In this study, we will examine an algorithm using only two successive frames. Therefore, temporal filter can not be used and hence only spatial filter is discussed. We assume that input images have 256×256 pixels and the number of resolution layers is 4. For each resolution l , the size N_l of a local region N_l , where depth is constant, has to be defined. The resolution l and the size N_l can be treated independently, but here, we simply define N_l so that intensity pattern has a slow slope in N_l . The set of N_l values and the corresponding spatial wavelengths, which components are extracted by an ideal band-pass filter, are shown in Table 1.

| layer number l | N_l [pixels] | spatial wavelength [pixels] |
|------------------|-------------------|--------------------------------|
| 1 | 32*32 | DC - 64 (DC - 4 cycles) |
| 2 | 16*16 | 64 - 32 (4 cycles - 8 cycles) |
| 3 | 8*8 | 32 - 16 (8 cycles - 16 cycles) |
| 4 | 8*8 | 16 - 8 (16 cycles - 32 cycles) |

Table 1. Decomposition parameters

6.2 Derivative filter

To get the accurate estimates of the spatial gradients of intensity, the choice of convolution kernels of derivative is important. Directly applying simple first-order differences produces poor estimates, especially in highly textured region.

By assuming no spatial alias occurs in images, the derivative of the continuous sampling function is the best kernel, but the resulting kernel needs to be quite large to estimate high accurate gradients. Hence, a lot of computer vision researchers have used sampled Gaussian derivatives that have better properties than simple differences, but are less computationally expensive than sampling function.

On the other hand, Farid and Simoncelli have proposed a simple design procedure for matched pairs of 1-D kernels, which consists of a interpolator and a differentiator, suitable for gradient estimation (Farid & Simoncelli, 1997). Let $\hat{B}(\mathbf{k})$ be the frequency domain representation of the interpolator, and $\hat{D}(\mathbf{k})$ be that of the differentiator. The kernel pairs determined by their procedure can have the following properties:

- 1. The derivative filters are good approximations to the derivative of the interpolator. This means that, for a derivative along the x-axis, $jk_x \hat{B}(\mathbf{k}) \approx \hat{D}(\mathbf{k})$ holds, where k_x is the component of the frequency coordinate in the x direction;
- 2. The interpolator is symmetric with $\hat{B}(0) = 1$;

3. Both kernels are separable, and hence the design problem is reduced to one dimensional, for computational efficiency and ease of design; and
4. The design algorithm includes a model for signal and noise statistics.

In this research, we use the kernels derived by this procedure, and concretely, the five-tap kernel $[-0.108415, -0.280353, 0, 0.280353, 0.108415]$ is applied to all the resolution layers as a differentiator.

As a temporal derivative, we compute simply the finite difference using two successive frames, hence the temporal derivative may be corrupted by large error.

6.3 Concrete procedures for M-step

At the M-step in the MAP-EM algorithm, we have to minimize Eq. 34 for each iteration. As described in Sec.5.2, the generalized MAP-EM algorithm can be adopted for estimating only the parameters values. However, the variances of the parameters are also required to be estimated, and therefore S-MAP-EM is performed. For the conventional S-MAP-EM algorithm denoted in Sec.5.3, it is assumed that the Q function is maximized at each M-step. If we use the generalized MAP-EM scheme, certain extension of S-MAP-EM has to be achieved. Hence, we will give up applying generalized MAP-EM and completely minimize Eq. 34 at the M-step.

Since Eq. 34 is not a quadratic form with respect to the parameter vector, we will use the steepest descent method to minimize it. We introduce the gradient vector, i.e. the steepest descent direction, of Eq. 34. By defining $\mathbf{m} \equiv [u_x, u_y, u_z, \mathbf{r}^T]^T$, partial derivative with respect to \mathbf{m} can be written as follows:

$$\frac{\partial J(\Theta)}{\partial \mathbf{m}} = \frac{2}{\sigma_1^2} \left(\frac{\partial \mathbf{m}_0}{\partial \mathbf{m}} \right)^T (\mathbf{A} \mathbf{m}_0 - \mathbf{b}) + 2 \tilde{\mathbf{V}}_m^{-1} (\mathbf{m} - \tilde{\mathbf{m}}). \quad (46)$$

The matrix \mathbf{A} and vector \mathbf{b} in Eq. 46 are shown in Sec. 10.1. Additionally, partial derivatives with respect to σ_0^2 and σ_1^2 can be introduced as follows:

$$\frac{\partial J(\Theta)}{\partial \sigma_0^2} = \sum_{i=1}^{M^{(l)}} \frac{1}{\mathbf{I}^{(l-1)^2} [\tilde{\sigma}_d^{2(l-1)}]_i + \sigma_0^2} - \sum_{i=1}^{M^{(l)}} \frac{\gamma_i}{\left(\mathbf{I}^{(l-1)^2} [\tilde{\sigma}_d^{2(l-1)}]_i + \sigma_0^2 \right)^2} + \frac{2(\sigma_0^2 - \tilde{\sigma}_0^2)}{\tilde{\sigma}_0^2}, \quad (47)$$

$$\frac{\partial J(\Theta)}{\partial \sigma_1^2} = \frac{N^{(l)} M^{(l)}}{\sigma_1^2} - \frac{\chi(\mathbf{m})}{\sigma_1^4} + \frac{2(\sigma_1^2 - \tilde{\sigma}_1^2)}{\tilde{\sigma}_1^2}, \quad (48)$$

where γ_i and $\chi(\mathbf{m})$ are defined as $J_d = \sum \gamma_i / (\mathbf{I}^{(l-1)^2} [\tilde{\sigma}_d^{2(l-1)}]_i + \sigma_0^2)$ and $J_f = \chi(\Theta) / \sigma_1^2$.

By evaluating the values of Eqs. 46, 47 and 48 at the current values of parameters, we can know the steepest descent direction and perform numerical search for the minimization parameters of $J(\Theta)$ defined in Eq. 34.

6.4 Parameter variance with complete data

In order to estimate the variance-covariance matrix of the parameters using the SEM scheme, we have to know the variance-covariance matrix of the complete data $\mathbf{V}_{c\hat{\Theta}}$ in Eqs. 39 and 41. We will compute analytically the 2nd derivative according to Eq. 40. Let $L(\Theta)$ be the conditional log likelihood function in Eq. 40 and $\Theta_0 \equiv [\mathbf{m}_0, \sigma_0^2, \sigma_1^2]^T$, and we can use $\partial L / \partial \Theta|_{\Theta=\hat{\Theta}} = \mathbf{0}$, hence the following equation can be derived

$$-\frac{\partial^2 L}{\partial \Theta \partial \Theta^T} \Big|_{\Theta=\hat{\Theta}} = -\left(\frac{\partial \Theta_0}{\partial \Theta}\right)^T \Big|_{\Theta=\hat{\Theta}} \left(\frac{\partial^2 L}{\partial \Theta_0 \partial \Theta_0^T}\right)_{\Theta=\hat{\Theta}} \left(\frac{\partial \Theta_0}{\partial \Theta}\right)_{\Theta=\hat{\Theta}}. \quad (49)$$

In the right hand side of Eq. 49, $\partial \Theta_0 / \partial \Theta$ is shown as follows:

$$\frac{\partial \Theta_0}{\partial \Theta} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -u_x / \sqrt{1-u_x^2-u_y^2} & -u_y / \sqrt{1-u_x^2-u_y^2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (50)$$

Additionally, $\partial^2 L / \partial \Theta_0 \partial \Theta_0^T$ in Eq. 49 consists of the components shown in Eq. A5. By evaluating Eqs. 50 and A5 at $\Theta = \hat{\Theta}$, we can know $-\partial^2 L / \partial \Theta \partial \Theta^T|_{\hat{\Theta}}$ from Eq. 49. On the other hand, log of the prior density $\ln p(\Theta|\{f_t^{(l)}\}, \dots, \{f_t^{(1)}\})$ is assumed to be

$$\ln p(\Theta|\{f_t^{(l)}\}, \dots, \{f_t^{(1)}\}) = -\frac{1}{2}(\mathbf{m} - \tilde{\mathbf{m}})^T \tilde{\mathbf{V}}_m^{-1}(\mathbf{m} - \tilde{\mathbf{m}}) - \frac{(\sigma_0^2 - \tilde{\sigma}_0^2)^2}{2\tilde{\sigma}_{\sigma_0}^2} - \frac{(\sigma_1^2 - \tilde{\sigma}_1^2)^2}{2\tilde{\sigma}_{\sigma_1}^2} + \text{Const.}, \quad (51)$$

hence,

$$-\frac{\partial^2}{\partial \Theta \partial \Theta^T} \ln p(\Theta|\{f_t^{(l)}\}, \dots, \{f_t^{(1)}\}) \Big|_{\Theta=\hat{\Theta}} = \begin{bmatrix} \tilde{\mathbf{V}}_m^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}^T & 1/\tilde{\sigma}_{\sigma_0}^2 & 0 \\ \mathbf{0}^T & 0 & 1/\tilde{\sigma}_{\sigma_1}^2 \end{bmatrix}. \quad (52)$$

Using Eqs. 49 and 52, $\mathbf{V}_{c\hat{\Theta}}$ can be computed as follows:

$$\mathbf{V}_{c\hat{\Theta}} = \left\{ - \left(\frac{\partial \Theta_0}{\partial \Theta} \right)^T_{\Theta=\hat{\Theta}} \left(\frac{\partial^2 L}{\partial \Theta_0 \partial \Theta_0^T} \right)_{\Theta=\hat{\Theta}} \left(\frac{\partial \Theta_0}{\partial \Theta} \right)_{\Theta=\hat{\Theta}} + \begin{bmatrix} \tilde{\mathbf{V}}_m^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}^T & 1/\tilde{\sigma}_{\sigma 0}^2 & 0 \\ \mathbf{0}^T & 0 & 1/\tilde{\sigma}_{\sigma 1}^2 \end{bmatrix} \right\}^{-1}. \quad (53)$$

7. Examples

To confirm the effectiveness of the proposed method, we conducted numerical experiments using artificial images. Figure 3(a) shows the first image, with 256×256 pixels generated by a computer graphics (CG) technique using the depth map shown in Fig. 3(b) and a random texture. The second successive image was generated at a different viewpoint, which was assumed to move with $\mathbf{u} = [0.1, 0.0, 0.0]^T$ and $\mathbf{r} = [0.0, 0.0, 0.0]^T$. In this situation, the theoretically calculated norm of the optical flow between the two successive images was approximately two pixels on average for the whole image. These images were decomposed into four layers with different resolutions in accordance with the method and the parameters described in Sec. 6.1. The decomposed images are shown in Fig. 4.

The estimated depth maps are shown in Fig. 5. As mentioned in Sec. 2.2, $\|\mathbf{u}\|$ and $|d|$ can not be uniquely determined, and hence the scale of the depth in Fig. 5 is adjusted so that the value of the estimated $\|\mathbf{u}\|$ can be regarded as a true value. The mesh size in Fig. 5 is denoted by N_l , and for example, $N_1 = 32 \times 32$ pixels. In the experiments, the variances of d and Θ in the priors to $l = 1$ were set to be sufficiently large. The result obtained using all the observed information corresponding to the four layers at the same time without BP is shown in Fig. 6. For this result, the local region size was 8×8 pixels. From these results, we can confirm that stable recovery of the depth map is achieved by the proposed method.

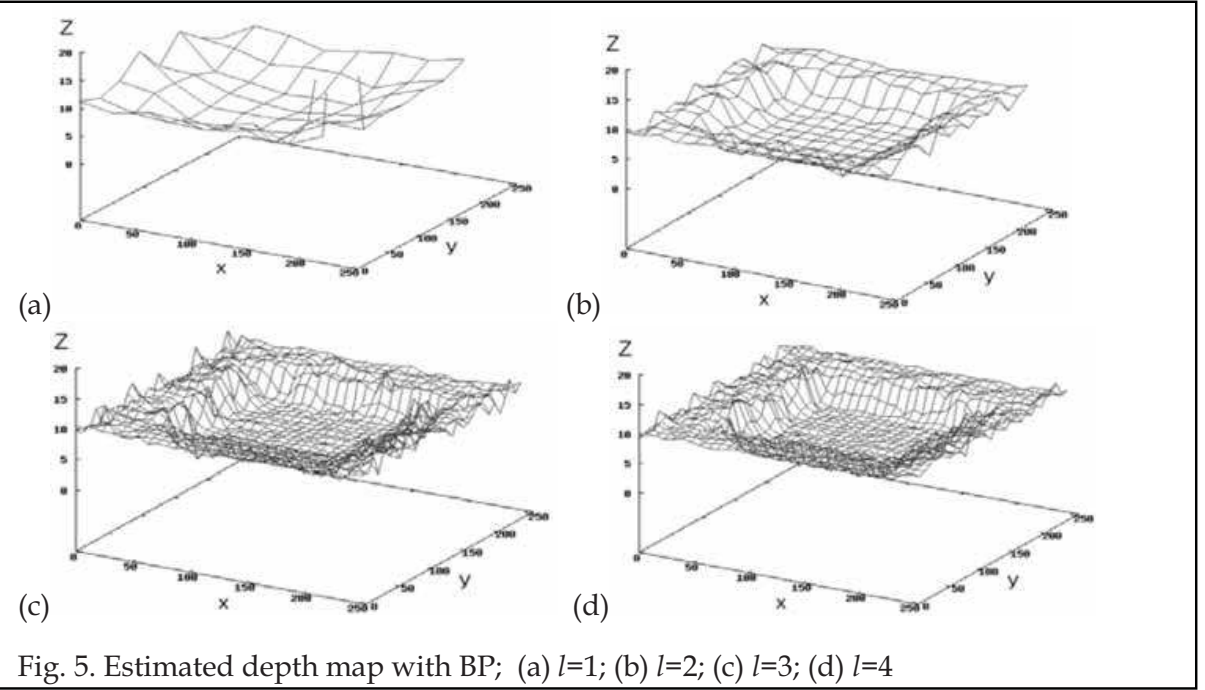
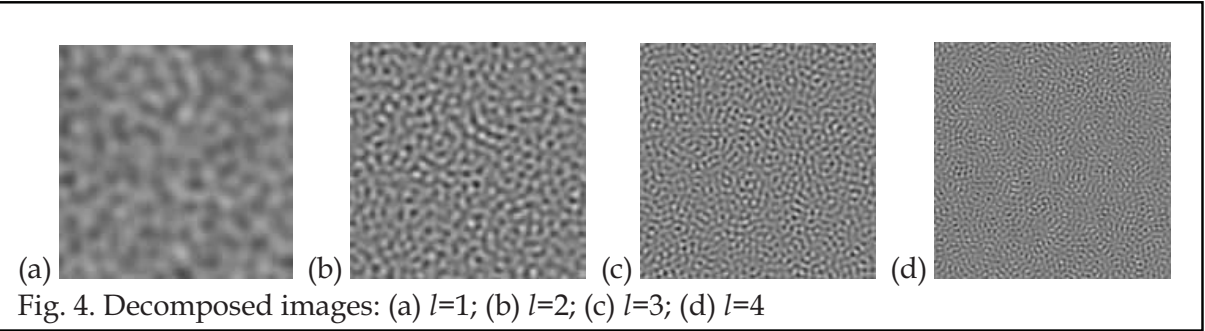
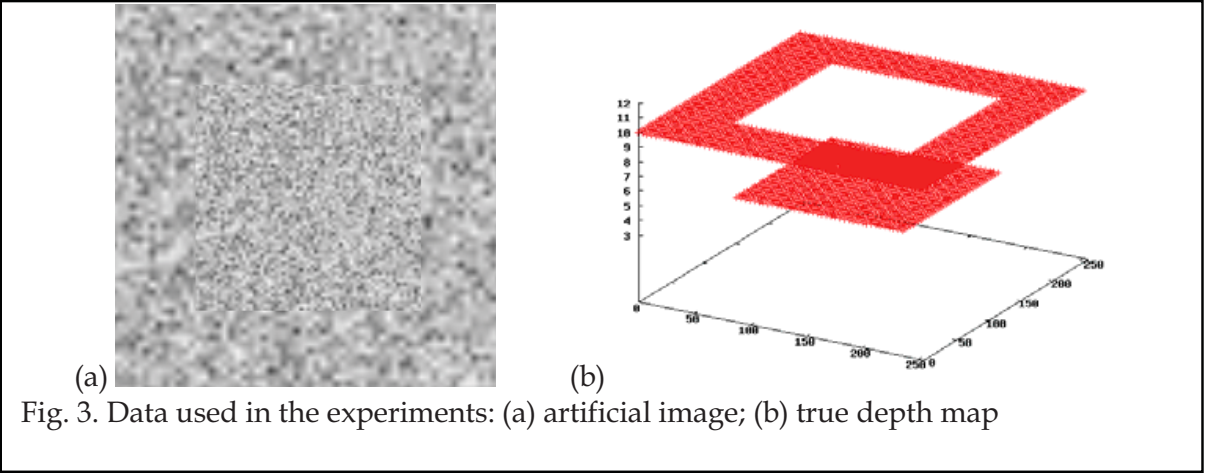
The above results were derived for noise-free images. Therefore, $n_1^{(l)}$ in Eq. 11 corresponds to the 1st approximation error of the gradient equation. We confirmed through experiments that, for the Gaussian image noise with a standard deviation of 5% with respect to the dynamic range of the image intensity, the proposed method has almost the same performance as that shown in Fig. 5. Additionally, we omitted BP of Θ , and found that the root mean square error (RMSE) of the depth is one and a half times larger than that using BP for Θ . This result is due to the estimation bias.

8. Conclusions

We introduced a scheme and an explicit algorithm for stably recovering object shape as a depth map. This scheme is based on the multi-scale Bayesian network and the approximate BP using the EM algorithm. Especially, in this study to estimate the variance-covariance matrix in a stable way, the Supplemented MAP-EM algorithm is applied. The effectiveness and the applicability of the proposed algorithm were shown through numerical examples. In the future, the performance for real image sequences needs to be examined, and a quantitative evaluation of the accuracy is required. Additionally, we are very interested in a temporal expansion of this scheme, and it has been considered in our current work.

9. Acknowledgments

This work has been supported by the research funds from Tokyo Metropolitan University. The authors are thankful to Dr. Todorka Alexandrova for her cooperation.



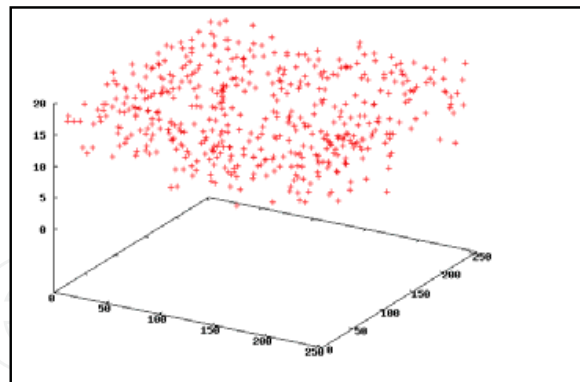


Fig. 6. Estimated depth map without BP

10. References

- Adiv, G. (1985). Determining three-dimensional motion and structure from optical flow generated by several moving objects, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 7, No. 4, pp. 384-401
- Brox, T., Bruhn, A., Papenberg, N., Weickert, J. (2004). High accuracy optical flow estimation based on a theory for warping, *Proceedings of ECCV*, Vol. 4, pp. 171-177
- Bruhn, A. & Weickert, J. (2005). Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods, *Int. J. Comput. Vision*, Vol. 61, No. 3, pp. 211-231
- Daniilidis, K. & Nagel, H.H. (1990). Analytical results on error sensitivity of motion estimation from two views, *Image and Vision Computing*, Vol. 8, pp. 297-303
- Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data, *J. Roy. Statist. Soc. B*, Vol. 39, pp. 1-38
- van Dyk, D.A., Meng, X.L., Rubin, D.B. (1995). Maximum likelihood estimation via the ECM algorithm: Computing the asymptotic variance, *Statistica Sinica*, Vol. 5, pp. 55-75
- Farid, H. & Simoncelli, E.P. (1997). Optimally rotation-equivariant directional derivative kernels, *Proceedings of Int. Conf. Computer Analysis of Image and Patterns*, pp. 207-214
- Farneback, G. (2001). Very high accuracy velocity estimation using orientation tensors, parametric motion, and simultaneous segmentation of the motion field, *Proceedings of ICCV*, pp. 171-177
- Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No. 6, pp. 721-741
- Han, M. & Kanade, T. (2002). A perspective factorization method for Euclidean reconstruction with uncalibrated cameras, *J. of Visualization and Computer Animation*, Vol. 13, No. 4, pp. 211-223
- Huang, T.S. & Faugeras, O.D. (1989). Some properties of the E matrix in two-view motion estimation, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, No. 12, pp. 1310-1312
- Huang, T.S. & Ho, H.T. (1999). A Kalman filter approach to direct depth estimation incorporating surface structure, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 21, No. 6, pp. 570-575
- Horn, B.K.P. & Schunk, B. (1981). Determining optical flow, *Artif. Intell.*, Vol. 17, pp. 185-203
- Horn, B.K.P. & Weldon Jr, E.J. (1988). Direct methods for recovering motion, *Int. J. Comput. Vision*, Vol. 2, pp. 51-76

- Kanatani, K. (1993). *Geometric Computation for Machine Vision*, Oxford University Press, Oxford
- Kanatani, K. (1996). *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier, Amsterdam
- Ke, Q. & Kanade, T. (2005). Robust L1 norm factorization in the presence of outliers and missing data by alternative convex programming, *Proceedings of CVPR*, pp. 739-746
- Kearney, J.K., Thompson, W.B., Boley, D.L. (1987). Optical flow estimation: An error analysis of gradient-based methods with local optimization, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 9, No. 2, pp. 229-244
- Longuet-Higgins, H.C. (1981). A computer algorithm for reconstructing a scene from two projections, *Nature*, Vol. 293, pp. 133-135
- Lucas, B.D. & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision, *Proceedings of Imaging Understanding Workshop*, pp. 121-130
- Maki, A., Watanabe, M., Wiles, C. (2002). Geotensity: Combining motion and lighting for 3D surface reconstruction, *Int. J. Comput. Vision*, Vol. 48, No. 2, pp. 75-90
- Matthies, L., Kanade, T., Szeliski, R. (1989). Kalman filter-based algorithm for estimating depth from image sequences, *Int. J. Comput. Vision*, Vol. 3, No. 3, pp. 209-238
- Maybank, S. (1990). Ambiguity in reconstruction from image correspondences, *Proceedings of ECCV*, pp. 177-186
- Meng, X.L., Rubin, D.B. (1991). Using EM to obtain asymptotic variance-covariance matrix: The SEM algorithm, *J. of the American Statist. Assoc.*, Vol. 86, No. 416, pp.899-909
- Nestares, O., Fleet, D.J., Heeger, D.J. (2000). Likelihood functions and confidence bounds for total-least-squares problems, *Proceedings of CVPR*, pp. 760-767
- Simoncelli, E.P. (1999). Bayesian multi-scale differential optical flow, In: *Handbook of Computer Vision and Applications*, Jähne, B., Haussecker, H, Geissler, P., (Eds.), pp. 397-422, Academic Press, San Diego
- Stein, G.P. & Shashua, A. (1997). Model-based brightness constraints: On direct estimation of structure and motion, *Proceedings of CVPR*, pp. 400-406
- Tagawa, N., Toriu, T., Endoh, T. (1993). Un-biased linear algorithm for recovering three-dimensional motion from optical flow, *IEICE Trans. Inf. & Syst.*, Vol. E76-D, No. 10, pp. 1263-1275
- Tagawa, N., Toriu, T., Endoh, T. (1994). An objective function for 3-D motion estimation from optical flow with lower error variance than maximum likelihood estimator, *Proceedings of IEEE Int. Conf. on Image Processing*, pp. 252-256
- Tagawa, N., Moriya, T. (1995). Computing 2-D motion field with multi-resolution images and cooperation of gradient-based and matching-based schemes, *IEICE Trans. Fundamentals*, Vol. E78-A, No. 6, pp. 685-692
- Tagawa, N., Toriu, T., Endoh, T. (1996). 3-D motion estimation from optical flow with low computational cost and small variance, *IEICE Trans. Inf. & Syst.*, Vol. E79-D, No. 3, pp. 230-241
- Tagawa, N., Kawaguchi, J., Naganuma, S., Okubo, K. (2008). Direct 3-D shape recovery from image sequence based on multi-scale Bayesian network, *Proceedings of ICPR*, CD
- Tsai, R.Y. & Huang, T.S. (1984). Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surface, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No. 6, pp. 13-27

Weiss, Y. & Fleet, D.J. (2001). Velocity likelihoods in biological and machine vision, In : *Probabilistic Models of the Brain : Perception and Neural Function*, Rao, R.P.N., Olshausen, B.A., Lewicki, M.S., Jahne, (Eds.), pp. 81-100, MIT Press, Cambridge

Zhuang, X., Huang, T.S., Ahuja, N., Haralick, R.M. (1988). A simplified linear optical flow-motion algorithm, *Comput. Vision, Graphics, Image Processing*, Vol. 42, pp. 334-344

11. Appendix

11.1 Definitions of matrix **A** and vector **b** in Eq. 46

The symmetric matrix **A** and vector **b** in Eq. 46 are explicitly defined as follows:

$$\begin{aligned}
 A_{11} &= \sum_{i=1}^{M^{(l)}} \alpha_i \sum_{N_i} f_x^{(l)2}, \quad A_{12} = \sum_{i=1}^{M^{(l)}} \alpha_i \sum_{N_i} f_x^{(l)} f_y^{(l)}, \quad A_{13} = - \sum_{i=1}^{M^{(l)}} \alpha_i \sum_{N_i} f_x^{(l)} (f_x^{(l)} x + f_y^{(l)} y), \quad A_{14} = - \sum_{i=1}^{M^{(l)}} \beta_i \sum_{N_i} f_x^{(l)} (f_x^{(l)} xy + f_y^{(l)} (1+y^2)), \\
 A_{15} &= \sum_{i=1}^{M^{(l)}} \beta_i \sum_{N_i} f_x^{(l)} (f_x^{(l)} (1+x^2) + f_y^{(l)} xy), \quad A_{16} = - \sum_{i=1}^{M^{(l)}} \beta_i \sum_{N_i} f_x^{(l)} (f_x^{(l)} y - f_y^{(l)} x), \quad A_{22} = \sum_{i=1}^{M^{(l)}} \alpha_i \sum_{N_i} f_y^{(l)2}, \quad A_{23} = - \sum_{i=1}^{M^{(l)}} \alpha_i \sum_{N_i} f_y^{(l)} (f_x^{(l)} x + f_y^{(l)} y), \\
 A_{24} &= - \sum_{i=1}^{M^{(l)}} \beta_i \sum_{N_i} f_y^{(l)} (f_x^{(l)} xy + f_y^{(l)} (1+y^2)), \quad A_{25} = \sum_{i=1}^{M^{(l)}} \beta_i \sum_{N_i} f_y^{(l)} (f_x^{(l)} (1+x^2) + f_y^{(l)} xy), \quad A_{26} = - \sum_{i=1}^{M^{(l)}} \beta_i \sum_{N_i} f_y^{(l)} (f_x^{(l)} y - f_y^{(l)} x), \\
 A_{33} &= \sum_{i=1}^{M^{(l)}} \alpha_i \sum_{N_i} (f_x^{(l)} x + f_y^{(l)} y)^2, \quad A_{34} = \sum_{i=1}^{M^{(l)}} \beta_i \sum_{N_i} (f_x^{(l)} x + f_y^{(l)} y) (f_x^{(l)} xy + f_y^{(l)} (1+y^2)), \quad A_{35} = - \sum_{i=1}^{M^{(l)}} \beta_i \sum_{N_i} (f_x^{(l)} x + f_y^{(l)} y) (f_x^{(l)} (1+x^2) + f_y^{(l)} xy), \\
 A_{36} &= \sum_{i=1}^{M^{(l)}} \beta_i \sum_{N_i} (f_x^{(l)} x + f_y^{(l)} y) (f_x^{(l)} y - f_y^{(l)} x), \quad A_{44} = \sum_{i=1}^{M^{(l)}} \sum_{N_i} (f_x^{(l)} xy + f_y^{(l)} (1+y^2))^2, \quad A_{45} = - \sum_{i=1}^{M^{(l)}} \sum_{N_i} (f_x^{(l)} xy + f_y^{(l)} (1+y^2)) (f_x^{(l)} (1+x^2) + f_y^{(l)} xy), \\
 A_{46} &= \sum_{i=1}^{M^{(l)}} \sum_{N_i} (f_x^{(l)} xy + f_y^{(l)} (1+y^2)) (f_x^{(l)} y - f_y^{(l)} x), \quad A_{55} = \sum_{i=1}^{M^{(l)}} \sum_{N_i} (f_x^{(l)} (1+x^2) + f_y^{(l)} xy)^2, \quad A_{56} = - \sum_{i=1}^{M^{(l)}} \sum_{N_i} (f_x^{(l)} (1+x^2) + f_y^{(l)} xy) (f_x^{(l)} y - f_y^{(l)} x), \\
 A_{66} &= \sum_{i=1}^{M^{(l)}} \sum_{N_i} (f_x^{(l)} y - f_y^{(l)} x)^2,
 \end{aligned} \tag{A1}$$

$$\begin{aligned}
 b_1 &= \sum_{i=1}^{M^{(l)}} \beta_i \sum_{N_i} f_t^{(l)} f_x^{(l)}, \quad b_2 = \sum_{i=1}^{M^{(l)}} \beta_i \sum_{N_i} f_t^{(l)} f_y^{(l)}, \quad b_3 = - \sum_{i=1}^{M^{(l)}} \beta_i \sum_{N_i} f_t^{(l)} (f_x^{(l)} x + f_y^{(l)} y), \quad b_4 = - \sum_{i=1}^{M^{(l)}} \sum_{N_i} f_t^{(l)} (f_x^{(l)} xy + f_y^{(l)} (1+y^2)), \\
 b_5 &= \sum_{i=1}^{M^{(l)}} \sum_{N_i} f_t^{(l)} (f_x^{(l)} (1+x^2) + f_y^{(l)} xy), \quad b_6 = - \sum_{i=1}^{M^{(l)}} \sum_{N_i} f_t^{(l)} (f_x^{(l)} y - f_y^{(l)} x),
 \end{aligned} \tag{A2}$$

where the coefficients α_i and β_i depending on the depth estimation are also defined

$$\alpha_i = \hat{d}_i^{(l)2} + \hat{\sigma}_{di}^{2(l)}, \quad \beta_i = \hat{d}_i^{(l)}, \tag{A3}$$

and the Jacobian $\partial \mathbf{m}_0 / \partial \mathbf{m}$ is represented as follows:

$$\frac{\partial \mathbf{m}_0}{\partial \mathbf{m}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -u_x / \sqrt{1-u_x^2-u_y^2} & -u_y / \sqrt{1-u_x^2-u_y^2} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{A4}$$

11.2 Definitions of matrix $\partial^2 L / \partial \Theta_0 \partial \Theta_0^T$ in Eq. 53

The components of $\partial^2 L / \partial \Theta_0 \partial \Theta_0^T$ can be written as follows:

$$\begin{aligned}
 \frac{\partial^2 L}{\partial u_x^2} &= -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} f_x^{(l)2} \tilde{d}_i^{(l)2}}{\sigma_1^2}, \quad \frac{\partial^2 L}{\partial u_y^2} = -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} f_y^{(l)2} \tilde{d}_i^{(l)2}}{\sigma_1^2}, \quad \frac{\partial^2 L}{\partial u_z^2} = -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} (f_x^{(l)} x + f_y^{(l)} y)^2 \tilde{d}_i^{(l)2}}{\sigma_1^2}, \\
 \frac{\partial^2 L}{\partial u_x \partial u_y} &= -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} f_x^{(l)} f_y^{(l)} \tilde{d}_i^{(l)2}}{\sigma_1^2}, \quad \frac{\partial^2 L}{\partial u_x \partial u_z} = -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} (f_x^{(l)} x + f_y^{(l)} y) f_x^{(l)} \tilde{d}_i^{(l)2}}{\sigma_1^2}, \quad \frac{\partial^2 L}{\partial u_y \partial u_z} = -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} (f_x^{(l)} x + f_y^{(l)} y) f_y^{(l)} \tilde{d}_i^{(l)2}}{\sigma_1^2}, \\
 \frac{\partial^2 L}{\partial r_x^2} &= -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} [f_x^{(l)} x y + f_y^{(l)} (1+y^2)]^2}{\sigma_1^2}, \quad \frac{\partial^2 L}{\partial r_y^2} = -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} [f_x^{(l)} (1+x^2) x y + f_y^{(l)} x y]^2}{\sigma_1^2}, \quad \frac{\partial^2 L}{\partial r_z^2} = -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} (f_x^{(l)} y - f_y^{(l)} x)^2}{\sigma_1^2}, \\
 \frac{\partial^2 L}{\partial r_x \partial r_y} &= -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} [f_x^{(l)} (1+x^2) + f_y^{(l)} x y] f_x^{(l)} x y + f_y^{(l)} (1+y^2)}{\sigma_1^2}, \quad \frac{\partial^2 L}{\partial r_x \partial r_z} = -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} [f_x^{(l)} x y + f_y^{(l)} (1+y^2)] f_x^{(l)} y - f_y^{(l)} x}{\sigma_1^2}, \\
 \frac{\partial^2 L}{\partial r_y \partial r_z} &= -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} [f_x^{(l)} (1+x^2) + f_y^{(l)} x y] f_x^{(l)} y - f_y^{(l)} x}{\sigma_1^2}, \quad \frac{\partial^2 L}{\partial u_x \partial r_x} = -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} [f_x^{(l)} x y + f_y^{(l)} (1+y^2)] f_x^{(l)} \tilde{d}_i^{(l)}}{\sigma_1^2}, \\
 \frac{\partial^2 L}{\partial u_x \partial r_y} &= -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} [f_x^{(l)} (1+x^2) + f_y^{(l)} x y] f_x^{(l)} \tilde{d}_i^{(l)}}{\sigma_1^2}, \quad \frac{\partial^2 L}{\partial u_x \partial r_z} = -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} (f_x^{(l)} y - f_y^{(l)} x) f_x^{(l)} \tilde{d}_i^{(l)}}{\sigma_1^2}, \\
 \frac{\partial^2 L}{\partial u_y \partial r_x} &= -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} [f_x^{(l)} x y + f_y^{(l)} (1+y^2)] f_y^{(l)} \tilde{d}_i^{(l)}}{\sigma_1^2}, \quad \frac{\partial^2 L}{\partial u_y \partial r_y} = -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} [f_x^{(l)} (1+x^2) + f_y^{(l)} x y] f_y^{(l)} \tilde{d}_i^{(l)}}{\sigma_1^2}, \\
 \frac{\partial^2 L}{\partial u_y \partial r_z} &= -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} (f_x^{(l)} y - f_y^{(l)} x) f_y^{(l)} \tilde{d}_i^{(l)}}{\sigma_1^2}, \quad \frac{\partial^2 L}{\partial u_z \partial r_x} = -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} [f_x^{(l)} x y + f_y^{(l)} (1+y^2)] f_x^{(l)} x + f_y^{(l)} y \tilde{d}_i^{(l)}}{\sigma_1^2}, \\
 \frac{\partial^2 L}{\partial u_z \partial r_y} &= -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} [f_x^{(l)} (1+x^2) + f_y^{(l)} x y] f_x^{(l)} x + f_y^{(l)} y \tilde{d}_i^{(l)}}{\sigma_1^2}, \quad \frac{\partial^2 L}{\partial u_z \partial r_z} = -\frac{\sum_{M^{(l)}} \sum_{N_i^{(l)}} (f_x^{(l)} y - f_y^{(l)} x) f_x^{(l)} x + f_y^{(l)} y \tilde{d}_i^{(l)}}{\sigma_1^2}, \\
 \frac{\partial^2 L}{\partial \sigma_0^2} &= -\sum_{M^{(l)}} \frac{(\tilde{d}_i^{(l)} - I^{(l-1)} [\tilde{d}^{(l-1)}])^2}{\left(I^{(l-1)2} [\sigma_d^{2(l-1)}] + \sigma_0^2 \right)^3} + \frac{1}{2} \sum_{M^{(l)}} \frac{1}{\left(I^{(l-1)2} [\sigma_d^{2(l-1)}] + \sigma_0^2 \right)^2}, \\
 \frac{\partial^2 L}{\partial \sigma_1^2} &= \frac{M^{(l)} N^{(l)}}{2(\sigma_1^2)^2} - \frac{1}{(\sigma_1^2)^3} \sum_{M^{(l)}} \sum_{N_i^{(l)}} \left[f_i^{(l)} - f_x^{(l)} \tilde{d}_i^{(l)} u_x - f_y^{(l)} \tilde{d}_i^{(l)} u_y + (f_x^{(l)} x + f_y^{(l)} y) \tilde{d}_i^{(l)} u_z + (f_x^{(l)} x y + f_y^{(l)} (1+y^2)) r_x - (f_x^{(l)} (1+x^2) + f_y^{(l)} x y) r_y + (f_x^{(l)} y - f_y^{(l)} x) r_z \right]^2, \\
 \frac{\partial^2 L}{\partial \sigma_0^2 \partial \sigma_1^2} &= \frac{\partial^2 L}{\partial \sigma_0^2 \partial u_x} = \frac{\partial^2 L}{\partial \sigma_0^2 \partial u_y} = \frac{\partial^2 L}{\partial \sigma_0^2 \partial u_z} = \frac{\partial^2 L}{\partial \sigma_0^2 \partial r_x} = \frac{\partial^2 L}{\partial \sigma_0^2 \partial r_y} = \frac{\partial^2 L}{\partial \sigma_0^2 \partial r_z} = 0, \\
 \frac{\partial^2 L}{\partial \sigma_1^2 \partial u_x} &= -\frac{1}{(\sigma_1^2)^2} \sum_{M^{(l)}} \sum_{N_i^{(l)}} \left[f_i^{(l)} - f_x^{(l)} \tilde{d}_i^{(l)} u_x - f_y^{(l)} \tilde{d}_i^{(l)} u_y + (f_x^{(l)} x + f_y^{(l)} y) \tilde{d}_i^{(l)} u_z + (f_x^{(l)} x y + f_y^{(l)} (1+y^2)) r_x - (f_x^{(l)} (1+x^2) + f_y^{(l)} x y) r_y + (f_x^{(l)} y - f_y^{(l)} x) r_z \right] f_x^{(l)} \tilde{d}_i^{(l)}, \\
 \frac{\partial^2 L}{\partial \sigma_1^2 \partial u_y} &= -\frac{1}{(\sigma_1^2)^2} \sum_{M^{(l)}} \sum_{N_i^{(l)}} \left[f_i^{(l)} - f_x^{(l)} \tilde{d}_i^{(l)} u_x - f_y^{(l)} \tilde{d}_i^{(l)} u_y + (f_x^{(l)} x + f_y^{(l)} y) \tilde{d}_i^{(l)} u_z + (f_x^{(l)} x y + f_y^{(l)} (1+y^2)) r_x - (f_x^{(l)} (1+x^2) + f_y^{(l)} x y) r_y + (f_x^{(l)} y - f_y^{(l)} x) r_z \right] f_y^{(l)} \tilde{d}_i^{(l)}, \\
 \frac{\partial^2 L}{\partial \sigma_1^2 \partial u_z} &= \frac{1}{(\sigma_1^2)^2} \sum_{M^{(l)}} \sum_{N_i^{(l)}} \left[f_i^{(l)} - f_x^{(l)} \tilde{d}_i^{(l)} u_x - f_y^{(l)} \tilde{d}_i^{(l)} u_y + (f_x^{(l)} x + f_y^{(l)} y) \tilde{d}_i^{(l)} u_z + (f_x^{(l)} x y + f_y^{(l)} (1+y^2)) r_x - (f_x^{(l)} (1+x^2) + f_y^{(l)} x y) r_y + (f_x^{(l)} y - f_y^{(l)} x) r_z \right] f_x^{(l)} x + f_y^{(l)} y \tilde{d}_i^{(l)}, \\
 \frac{\partial^2 L}{\partial \sigma_1^2 \partial r_x} &= \frac{1}{(\sigma_1^2)^2} \sum_{M^{(l)}} \sum_{N_i^{(l)}} \left[f_i^{(l)} - f_x^{(l)} \tilde{d}_i^{(l)} u_x - f_y^{(l)} \tilde{d}_i^{(l)} u_y + (f_x^{(l)} x + f_y^{(l)} y) \tilde{d}_i^{(l)} u_z + (f_x^{(l)} x y + f_y^{(l)} (1+y^2)) r_x - (f_x^{(l)} (1+x^2) + f_y^{(l)} x y) r_y + (f_x^{(l)} y - f_y^{(l)} x) r_z \right] (f_x^{(l)} x y + f_y^{(l)} (1+y^2)), \\
 \frac{\partial^2 L}{\partial \sigma_1^2 \partial r_y} &= -\frac{1}{(\sigma_1^2)^2} \sum_{M^{(l)}} \sum_{N_i^{(l)}} \left[f_i^{(l)} - f_x^{(l)} \tilde{d}_i^{(l)} u_x - f_y^{(l)} \tilde{d}_i^{(l)} u_y + (f_x^{(l)} x + f_y^{(l)} y) \tilde{d}_i^{(l)} u_z + (f_x^{(l)} x y + f_y^{(l)} (1+y^2)) r_x - (f_x^{(l)} (1+x^2) + f_y^{(l)} x y) r_y + (f_x^{(l)} y - f_y^{(l)} x) r_z \right] (f_x^{(l)} (1+x^2) + f_y^{(l)} x y), \\
 \frac{\partial^2 L}{\partial \sigma_1^2 \partial r_z} &= \frac{1}{(\sigma_1^2)^2} \sum_{M^{(l)}} \sum_{N_i^{(l)}} \left[f_i^{(l)} - f_x^{(l)} \tilde{d}_i^{(l)} u_x - f_y^{(l)} \tilde{d}_i^{(l)} u_y + (f_x^{(l)} x + f_y^{(l)} y) \tilde{d}_i^{(l)} u_z + (f_x^{(l)} x y + f_y^{(l)} (1+y^2)) r_x - (f_x^{(l)} (1+x^2) + f_y^{(l)} x y) r_y + (f_x^{(l)} y - f_y^{(l)} x) r_z \right] (f_x^{(l)} y - f_y^{(l)} x),
 \end{aligned}$$

(A5)



Pattern Recognition

Edited by Peng-Yeng Yin

ISBN 978-953-307-014-8

Hard cover, 568 pages

Publisher InTech

Published online 01, October, 2009

Published in print edition October, 2009

For more than 40 years, pattern recognition approaches are continually improving and have been used in an increasing number of areas with great success. This book discloses recent advances and new ideas in approaches and applications for pattern recognition. The 30 chapters selected in this book cover the major topics in pattern recognition. These chapters propose state-of-the-art approaches and cutting-edge research results. I could not thank enough to the contributions of the authors. This book would not have been possible without their support.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Norio Tagawa and Shoichi Naganuma (2009). Structure and Motion from Image Sequences Based on Multi-Scale Bayesian Network, Pattern Recognition, Peng-Yeng Yin (Ed.), ISBN: 978-953-307-014-8, InTech, Available from: <http://www.intechopen.com/books/pattern-recognition/structure-and-motion-from-image-sequences-based-on-multi-scale-bayesian-network>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2009 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen